Health Care Delivery

Diagnosis Cluster Frequency in a Community-Based Family Practice Residency Program
Comparison With Large Ambulatory Data Sets

CHARLES L. SHEAR, DrPH, and ERIC M. WALL, MD, MPH, San Bernardino, California

Ambulatory encounters in a community-based family practice residency program were analyzed using diagnosis clusters. During an 18-month period (July 1982 through December 1983), demographic information and clinical diagnoses for 44,453 successive patient visits were collected and stored in a computerized data base. The 30 most frequent diagnosis clusters accounted for 70% of all recorded clinical diagnoses. Comparison with NAMCS, USC-MAMP (Western Region) and Virginia studies revealed a younger, more indigent population with a higher frequency of visits for hypertension, prenatal and postnatal care, diabetes, chronic respiratory illness and congestive heart failure. Consistent with the other large ambulatory data sets, the general medical examination, hypertension and acute upper respiratory conditions were the most frequent diagnosis clusters. Differences with other reported data sets reflected the site-specific demographic characteristics of patients and providers, regional and environmental influences on the incidence of specific disease states and the relative abundance of other subspecialist physicians. Such local or regional data bases not only provide valuable information as to clinical content but also may help in identifying previously unrecognized health problems.

Analysis of the clustering of clinical diagnoses according to similar pathophysiologic conditions is a powerful tool to examine the content of ambulatory care. Such clusters purportedly elicit similar cognitive responses in the process of clinical decision making. In their landmark article, Rosenblatt and co-workers[1] have outlined a series of 92 separate diagnosis clusters that (1) encompass the majority of discrete diagnostic entities used in ambulatory practice, (2) decrease the idiosyncratic diagnosis labeling patterns of individual physicians and (3) conform to existing diagnostic classifications used currently in office-based practice.[2]

This technique has thus far been applied to large regional and national data sets that were collected by others and that reflect the content of office-based practice by general and family physicians in the 1970s. The National Ambulatory Medical Care Survey (NAMCS) is an ongoing effort of the National Center for Health Statistics which randomly surveys approximately 1,000 office-based physicians annually. Participating physicians are asked to complete a log-diary of patient encounters during a one-week period at some time from July through October.[3] The Medical Activities and Manpower Project at the University of Southern California (USC-MAMP) study utilized a random sample of physicians identified from the American Medical Association's physician masterfile. A log-diary of each patient encounter during a three-day period was recorded during the summer and fall of 1977.[4] The Virginia Study reported diagnostic data for clinical encounters made between 1973 and 1975 for 36 family physicians and 82 family practice residents practicing in the state of Virginia.[5]

ABBREVIATIONS USED IN TEXT

COPD = chronic obstructive pulmonary disease
NAMCS = National Ambulatory Medical Care Survey
SBCMC = San Bernardino County Medical Center
USC-MAMP = University of Southern California—
Medical Activities and Manpower Project

This study represents an analysis of ambulatory encounters using this technique in a community-based family practice residency program. The goals of this study were to analyze the experience of the first 18 months of an ongoing data base of ambulatory encounters, to validate the usefulness of the technique of diagnosis clusters and to compare the results with those recently reported—keeping in mind that such would reflect only one contemporary training setting for future family physicians.

Methods

The Department of Family Practice of the San Bernardino County Medical Center (SBCMC) presently operates four Family Health Centers in the urban area of San Bernardino County, which is located in the eastern section of the Los Angeles basin. In 1980 the greater San Bernardino metropolitan area had a population of 340,000. Its demographic profile is similar to that of California as a whole, with 82% of the population being white (race) and 19% being of Hispanic origin (ethnicity). Approximately 27% of the population has not completed high school and the median family income is about $20,000. The health status of the county is somewhat lower than that of California as a whole with an age-adjusted death rate of 8.7 per 1,000.

The four Family Health Centers provide the training for 52 family practice residents and the ambulatory care for approximately 32,000 patient visits per year. On July 1, 1982, a computerized system containing clinical and demographic information was initiated. This system is divided into two main data bases. One maintains family-based demographic information on "active patients" (those having made at least one visit to a health center in the preceding 18 months). For each member of a household, the data base includes primary physician name, patient name, age, sex, family relationship code (as indicated on a standard genogram), medical record number and family chart number. The source of information for this data base is registration forms completed by patients at the initial visit to the health center. A second data base maintains diagnostic information on ambulatory encounters at the Family Health Centers and includes the date of the visit as well as the primary and secondary diagnosis for that visit (ICD-9-CM). The source of this information is encounter forms completed by resident and faculty providers. Diagnosis codes (ICD-9-CM) are placed on the encounter forms by clerical personnel and are then verified by the Division of Research before data entry.

Data are entered by trained personnel into a microcomputer with numerous edit traps to improve the accuracy of input information. Periodically, the collected information is sent by telephone to a Harris mainframe computer at Loma Linda University for remote processing by SBCMC personnel. Information on resident patient panels and diagnosis frequency is then regularly distributed to providers and administrative personnel. In addition, when the two data sets are merged, longitudinal information on family utilization of the health centers and diagnostic trends within families can be obtained.

The present report describes the diagnosis clusters for all four Family Health Centers from July 1, 1982, to December 31, 1983. Clusters were constructed from individual ICD-9-CM diagnoses using the method of Schneeweis and co-

TABLE 1.—Demographic Characteristics of
Patient Populations for SBCMC Data Base, NAMCS
and USC-MAMP (%)

Age/Sex Category	SBCMC* (1982-1983)	NAMCS† (1977-1978)	USC-MAMP‡ (1977)
Sex			
Female—all	62.3	59.5	58.4
Male—all	37.7	40.5	41.6
Age (years)			
<17	38.7	18.1	19.1
17-44	38.2	39.1	39.9
45-64	13.8	25.5	24.1
65+	9.2	17.3	16.9

SBCMC=San Bernardino County Medical Center, NAMCS=National Ambulatory Medical Care Survey, USC-MAMP=University of Southern California-Medical Activities and Manpower Project

*Based on age-sex register of 10,733 patients.
†Based on 9,164 outpatient encounters.
‡Based on 38,466 outpatient encounters.

TABLE 2.—Frequency of Diagnosis Clusters in the SBCMC
Data Base (July 1982-December 1983)

		Frequency	
Rank	Cluster	N	(Cumulative %)
1	Gen. medical examination	7,194	(15.3)
2	Hypertension	3,783	(23.3)
3	Prenatal/postnatal care	3,555	(30.9)
4	Acute upper respiratory tract infection	2,661	(36.5)
5	Diabetes mellitus	2,311	(41.4)
6	Otitis media	1,618	(44.8)
7	Depression/anxiety	1,100	(47.2)
8	Asthma	968	(49.2)
9	Emphysema, COPD	900	(51.1)
10	Urinary tract infection	793	(52.8)
11	Soft tissue injury	698	(54.3)
12	Degenerative joint disease	636	(55.7)
13	Ischemic heart disease	622	(57.0)
14	Congestive heart failure	619	(58.2)
15	Abdominal pain	515	(59.4)
16	Nonfungal skin infection	512	(60.5)
17	Vaginitis/vulvitis	510	(61.6)
18	Contraception	377	(62.4)
19	Headaches	374	(63.2)
20	Peptic diseases	367	(63.9)
21	Acute lower respiratory tract infection	334	(64.6)
22	Fibrositis, myalgia	318	(65.3)
23	Thyroid disease	306	(66.0)
24	Dermatitis, eczema	305	(66.6)
25	Conjunctivitis, keratitis	302	(67.3)
26	Rheumatoid disease	278	(67.9)
27	Chest pain	260	(68.4)
28	Menstrual disorders	255	(68.9)
29	Sprains, strains	243	(69.5)
30	Seizure disorder	235	(70.0)

COPD=chronic obstructive pulmonary disease

workers.[2] Only primary diagnoses were used in the construction of these clusters. These represent 95% of all diagnoses recorded during this time.

Results

In all, 10,733 persons were identified as active patients during the period July 1, 1982, through December 31, 1983. Approximately 60% of these were female. Nearly 40% of the practice population was under 17 years of age with 10% over age 65; 51% of patients were white, 32% Hispanic and 13% black. As seen in Table 1 this population is considerably younger than those of the NAMCS and USC-MAMP studies. More than 85% of the SBCMC population receive Medi-Cal or Medicare assistance.

The ten most frequent diagnosis clusters represented 52.8% of the 44,453 primary diagnoses made during the study period with the top 30 clusters accounting for 70% of all diagnoses (Table 2). The general medical examination, hypertension and prenatal and postnatal care were the most frequent clusters accounting for approximately a third of all diagnoses made.

Comparison with other studies (see Table 3) indicates a higher frequency of hypertension, prenatal and postnatal care, diabetes mellitus, otitis media, asthma, emphysema, COPD (chronic obstructive pulmonary disease)-bronchitis, congestive heart failure and contraception in the SBCMC population. Notably less frequent were acute orthopedic injuries such as sprains, strains, soft tissue injury, fractures and dislocations. Rhinitis, sinusitis and acute lower respiratory tract infections were much less frequent than reported in other studies. Finally, obesity, infectious diarrhea/gastroenteritis

and medical/surgical aftercare were significantly lower in frequency in the SBCMC data base.

Discussion

The results of this study confirm the ability of diagnosis clusters to effectively reduce the quantity of diagnostic information into a more manageable and meaningful number. In the SBCMC data base, 70% of all diagnoses were captured in the 30 most frequent clusters. This is comparable to the experience of Schneeweis and associates.[2] Independent of the method of data collection, three diagnosis clusters have been found consistently in all reported ambulatory data sets. These include the general medical examination, hypertension and acute upper respiratory tract conditions. These three will account for approximately 25% to 30% of all diagnoses.

The diverse spectrum of clinical diagnoses encountered in family practice is also seen in this study. Those factors influencing this spectrum include the demographic characteristics of patients and physicians, the incidence of specific disease states, the relative supply of other physician specialists and the focus or setting of care.[1] The SBCMC data reported here are from one residency training setting where the majority of physicians are 25 to 32 years of age. This may have influenced the age distribution of the patient population which has twice as many persons younger than 17 years as the NAMCS and USC-MAMP studies. This age difference may explain in part the high relative frequency of prenatal and postnatal care, otitis media and contraception.

The high relative frequency of hypertension, diabetes and congestive heart failure at first seems inconsistent with this younger population. On the other hand, socioeconomic and

TABLE 3.—Most Frequent Diagnosis Clusters Encountered at SBCMC and Comparison With USC-MAMP, NAMCS and Virginia Studies

Diagnosis Cluster	SBCMC* Rank	SBCMC* %	USC-MAMP† Rank	USC-MAMP† %	NAMCS‡ Rank	NAMCS‡ %	Virginia§ Rank	Virginia§ %
General medical examination	1	(15.3)	1	(12.2)	1	(9.8)	1	(11.7)
Hypertension	2	(7.9)	3	(4.8)	3	(6.8)	3	(5.8)
Prenatal and postnatal care	3	(7.5)	6	(3.3)	7	(2.9)	13	(1.5)
Acute upper respiratory tract infection ...	4	(5.6)	2	(7.5)	2	(9.3)	2	(8.4)
Diabetes mellitus	5	(4.9)	9	(1.9)	9	(2.4)	8	(2.4)
Otitis media	6	(3.4)	21	(1.1)	17	(1.4)	11	(1.9)
Depression/anxiety	7	(2.3)	7	(2.7)	10	(2.3)	6	(3.8)
Asthma	8	(2.1)	>25		29	(0.7)	>30	
Emphysema, COPD	9	(1.9)	20	(1.3)	>30		>30	
Urinary tract infection	10	(1.7)	15	(1.5)	16	(1.7)	17	(1.2)
Soft tissue injury	11	(1.5)	5	(4.2)	4	(4.2)	4	(4.0)
Degenerative joint disease	12	(1.4)	10	(1.8)	13	(2.1)	15	(1.4)
Ischemic heart disease	13	(1.3)	14	(1.5)	8	(2.5)	9	(2.2)
Congestive heart failure	14	(1.3)	>25		>30		>30	
Abdominal pain	15	(1.1)	>25		>30		>30	
Nonfungal skin infections	16	(1.1)	24	(0.9)	20	(1.2)	12	(1.7)
Vaginitis/vulvitis	17	(1.1)	11	(1.8)	23	(1.1)	16	(1.2)
Contraception	18	(0.8)	>25		>30		>30	
Headaches	19	(0.8)	23	(1.1)	27	(0.7)	24	(0.8)
Peptic diseases	20	(0.8)	22	(1.1)	21	(1.2)	22	(0.9)

SBCMC=San Bernardino County Medical Center, USC-MAMP=University of Southern California-Medical Activities and Manpower Project, NAMCS=National Ambulatory Medical Care Survey, COPD=chronic obstructive pulmonary disease

*Based on 44,453 outpatient encounters (1982-1983).
†Based on Western Region encounters only (1977).
‡Based on 9,164 outpatient encounters (1977-1978).
§Based on 526,196 encounters (1973-1975).

ethnic factors may well be influential in determining the content of practice in this setting. Most of these patients receive medical entitlement benefits and approximately a third are of Hispanic descent. It is significant that this latter group has been found to be at particularly high risk for diabetes mellitus.[6]

A higher frequency of obesity would be expected in view of the increased visits reported for diabetes and heart disease. No such association was found in this study, perhaps because physicians recorded this as a secondary diagnosis. If this is true, it suggests that this training setting fails to adequately address the importance of this problem.

The influence of a relative abundance of other specialists may also explain some of the results. The low frequency of acute orthopedic injuries, for example, may be a result of the presence of appropriate subspecialty services and an emergency room department available in the same institution.

Finally, the influence of environmental conditions on clinical content is strongly supported by the SBCMC data. The very high visit rates for emphysema, COPD, bronchitis and asthma may be explained by the poor air quality of the Los Angeles basin which characteristically worsens from May through October. Preliminary analysis of cluster frequency by month lends support to this hypothesis. Further investigation of this phenomenon is planned at a later date.

In conclusion, the results of this study put into perspective the reports of national versus regional encounter data sets.

While large national data bases do give an accurate estimate of national trends, the application of these findings to local practices and training settings must be viewed with some skepticism. Local influences of training, physician supply, environment and characteristics of patients and providers may dramatically determine the clinical content of such encounters. Consequently, it is recommended that similar data bases be developed locally or regionally so that family practice residency programs can more accurately determine the proper balance of clinical content appropriate for their needs and goals. The availability of such information to community physicians may also serve an important function in identifying local health problems that otherwise would go unnoticed.

REFERENCES

1. Rosenblatt RA, Cherkin DC, Schneeweis R, et al: The structure and content of family practice: Current status and future trends. J Fam Pract 1982; 15:681-722

2. Schneeweis R, Rosenblatt RA, Cherkin DC, et al: Diagnosis clusters: A new tool for analyzing the content of ambulatory medical care. Med Care 1983 Jan; 21:105-122

3. The national ambulatory medical care survey: Background and methodology. Vital and Health Statistics, Series 2, No. 61. DHEW publication No. (HRA) 74-1335. National Center for Health Statistics, Hyattsville, Md, Government Printing Office, 1974

4. Mendenhall RC, Lloyd JS, Repicky PA, et al: A national study of medical and surgical specialties: II. Description of the survey instrument. JAMA 1978 Sep 8; 240:1160-1168

5. Marsland DW, Wood M, Mayo F: A data bank for patient care, curriculum, and research in family practice: 526,196 patient problems. J Fam Pract 1976 Jan/Feb; 3:25-28

6. Stern MP, Gaskill SP, Allen CR Jr, et al: Cardiovascular risk factors in Mexican-Americans in Laredo, Texas: I. Prevalence of overweight and diabetes and distributions of serum lipids. Am J Epidemiol 1981 May; 113:546-555